

## 8 Martingale convergence and Azuma's inequality

### 8.1 Setup: what is a martingale?

#### Definition 8.1

A **martingale** is a sequence of random variables  $Z_0, Z_1, \dots$ , such that for every  $n$ ,  $\mathbb{E}|Z_n| < \infty$  (this is a technical assumption), and

$$\mathbb{E}[Z_{n+1}|Z_0, \dots, Z_n] = Z_n.$$

This comes up in a lot of different ways:

#### Example 8.2

Consider a random walk  $X_1, X_2, \dots$  of independent steps, each with mean 0. Then we can define the martingale

$$Z_n = \sum_{i \leq n} X_i,$$

which fits the definition because we always expect our average position after step  $n + 1$  to be the same as where we just were after step  $n$ .

#### Example 8.3 (Betting strategy)

Let's say we go to a casino, and all bets are "fair" (have expectation 0). For example, we may bet on fair odds against a coin flip. Our strategy can adapt over time based on the outcomes: let  $Z_n$  be our balance after  $n$  rounds. This is still a martingale!

This is more general than just a random walk, because now we don't need the steps to be independent.

#### Example 8.4

Let's say my goal is to win 1 dollar. I adapt the following strategy:

- Bet a dollar; if I win, stop.
- Otherwise, double the wager and repeat.

This is a martingale, because all betting strategies are martingales. With probability 1, we must always win at some point, so we end up with 1 dollar at the end! This sounds like free money, but we have a finite amount of money (so this would never occur in real life).

**Remark.** This is called the "martingale betting strategy," and it's where the name comes from!

#### Definition 8.5 (Doob or exposure martingale)

Suppose we have some (not necessarily independent) random variables  $X_1, \dots, X_n$ , and we have a function  $f(x_1, \dots, x_n)$ . Then let

$$Z_i = \mathbb{E}[f(x_1, \dots, x_n) | X_1, \dots, X_i].$$

Basically, we “expose” the first  $i$  outputs to create  $Z_i$ . It’s good to check that this is actually a martingale: show that

$$\mathbb{E}[Z_i | Z_0, \dots, Z_{i-1}] = Z_{i-1}.$$

Note that  $f$  may also be some random variable: for example, it could be the chromatic number of the graph, and  $X_i$  are indicator variables of the edges. Then  $Z_0$  is  $\mathbb{E}[f]$ ,  $Z_1$  is a revised mean after we learn about the status of an edge, and so on. This is called an **edge-exposure martingale**.

Let’s discuss that more explicitly: we reveal the edges of  $G(n, p)$  one at a time. For example, let’s say we want  $\chi(G(3, \frac{1}{2}))$ . There are eight possible graphs, with equal probabilities, and six of them have chromatic number 2, one has chromatic number 3, and one has chromatic number 1. The average is  $Z_0 = 2$ .

Now, the chromatic number is either 2.25 or 1.75, depending on whether the bottom edge is present or not. This average is 2, and then we can keep going:  $Z_2$  is either 2.5 or 2 if  $Z_1 = 2.25$ , and 2 or 1.5 if  $Z_1 = 1.75$ . The idea is that each  $Z_{n+1}$ ’s expected value is dependent on the previous mean.

Alternatively, we can have a **vertex-exposure martingale**: at the  $i$ th step, expose all edges  $(j, i)$  with  $j < i$ . So there are different ways of constructing this martingale, and which one to use depends on the application!

## 8.2 Azuma’s inequality

Why are martingales useful? Here’s an important inequality that’s actually not too hard to prove:

### Theorem 8.6 (Azuma’s inequality)

Given a martingale  $Z_0, \dots, Z_n$  with bounded differences

$$|Z_i - Z_{i-1}| \leq 1 \forall i \in [n],$$

we have a tail bound for all  $\lambda$ :

$$\Pr(Z_n - Z_0 \geq \lambda\sqrt{n}) \leq e^{-\lambda^2/2}.$$

More generally, though, if we have  $|Z_i - Z_{i-1}| \leq c_i$  for all  $i \in [n]$ , then for all  $a > 0$ ,

$$\Pr(Z_n - Z_0 \geq a) \leq \exp\left(\frac{-a^2}{2\sum_{i=1}^n c_i^2}\right).$$

(This is sometimes also known as Azuma-Hoeffding.) We’ve seen this before from our discussion of Chernoff bounds, which is a special case by making the martingale a sum of independent random variables! This is not a coincidence - we’ll notice similarities in the proof.

This theorem is useful when none of the martingale steps have big differences. It is generally more difficult to prove any sort of concentration when we don’t have bounded differences in our martingale, though.

*Proof.* We can shift the martingale so that  $Z_0 = 0$ . Let  $X_i = Z_i - Z_{i-1}$  be the martingale differences: these  $X_i$ s do not need to be independent, but they must always have mean 0.

### Lemma 8.7

If  $X$  is a random variable with  $\mathbb{E}[X] = 0$  and  $|x| \leq c$ , then

$$\mathbb{E}[e^X] \leq \frac{e^c + e^{-c}}{2} \leq e^{c^2/2}$$

by looking at Taylor expansion and comparing coefficients.

*Proof of lemma.* Basically, we maximize  $e^x$  by having a  $\pm c$  Bernoulli variable - this is true because of convexity! Specifically, we can upper-bound  $e^x$  by the line connecting the endpoints:

$$e^x \leq \frac{e^c + e^{-c}}{2} + \frac{e^c - e^{-c}}{2c}x,$$

and now take expectations of the statement when  $x = X$ . □

So now let  $t \geq 0$ , and consider the moment generating function  $\mathbb{E}[e^{tZ_n}]$ . We'll split this up as

$$\mathbb{E}[e^{tZ_n}] = \mathbb{E}[e^{t(X_n + Z_{n-1})}] = \mathbb{E}[\mathbb{E}[e^{tX_n} | Z_{n-1}] e^{tZ_{n-1}}]$$

By definition, the inner expectation is the moment generating function of a mean-zero random variable bounded by  $t c_n$ , and thus

$$\mathbb{E}[e^{tZ_n}] \leq e^{t^2 c_n^2 / 2} \mathbb{E}[e^{tZ_{n-1}}].$$

Repeating this calculation or using induction, we find that the expectation of  $e^{tZ_n}$  is bounded by

$$\leq \exp\left[\frac{t^2}{2}(c_n^2 + c_{n-1}^2 + \dots + c_1^2)\right].$$

To finish the proof, we repeat the logic of the Chernoff bound proof: by Markov's inequality on the moment generating function,

$$\Pr(Z_n \geq a) \leq e^{-ta} \mathbb{E}[e^{tZ_n}] \leq e^{-ta + t^2(c_1^2 + \dots + c_n^2)/2}.$$

We can now set  $t$  to be whatever want: taking  $t = \frac{a}{\sum_i c_i^2}$  yields the result. □

The main difference from Chernoff is that we do one step at a time, and this crucially requires that we have bounded differences. We can also get a lower tail for  $Z_n$  in the exact same way, and putting these together yields the following:

### Corollary 8.8

Let  $Z_n$  be a martingale where  $|Z_i - Z_{i-1}| \leq c_i$  for all  $i \in [n]$ , as in Theorem 8.6. Then for all  $a > 0$ ,

$$\Pr(|Z_n - Z_0| \geq a) \leq 2 \exp\left(\frac{-a^2}{2 \sum_i c_i^2}\right).$$

Basically, we can't walk very far in either direction in a martingale with an interval of  $\sqrt{n}$ , even when our choices can depend on previous events.

## 8.3 Basic applications of this inequality

The most common way Azuma is used is to show concentration for Lipschitz functions (on a domain of many variables).

### Theorem 8.9

Consider a function

$$f : \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n \rightarrow \mathbb{R}$$

such that  $|f(x) - f(y)| \leq 1$  whenever  $x$  and  $y$  are vectors that differ in exactly 1 coordinate. (This is known as being **1-Lipschitz** with respect to Hamming distance.) Then if  $Z = f(X_1, \dots, X_n)$  is a function of **independent random variables**  $X_i \in \Omega_i$ , we have high concentration:

$$\Pr(Z - \mathbb{E}[Z] \geq \lambda\sqrt{n}) \leq e^{-\lambda^2/2}.$$

*Proof.* Consider the Doob martingale

$$Z_i = \mathbb{E}[Z | X_1, \dots, X_i].$$

Note that  $|Z_i - Z_{i-1}| \leq 1$ , because revealing 1 coordinate cannot change the value of our function by more than 1. But now  $Z_0$  is the expected value of our original function  $Z$ , since we have no additional information: thus,  $Z_0 = \mathbb{E}[Z]$ . Meanwhile,  $Z_n$  means we have all information about our function, so this is just  $Z$ . Plugging these into the Azuma inequality yields the result.  $\square$

It's important to note that the  $|Z_i - Z_{i-1}| \leq 1$  step only works if we have independence between our random variables - that step is a bit subtle.

### Example 8.10 (Coupon collecting)

Let's say we want to collect an entire stack of coupons: we sample  $s_1, \dots, s_n \in [n]$ . Can we describe  $X$ , the number of missed coupons?

Explicitly, we can write out

$$X = |[n] \setminus \{s_1, \dots, s_n\}|.$$

It's not hard to calculate the expected value of  $X$ : by linearity of expectation, each coupon is missed with probability  $(1 - \frac{1}{n})^n$ , so

$$\mathbb{E}[X] = n \left(1 - \frac{1}{n}\right)^n.$$

This value is between  $\frac{n-1}{e}$  and  $\frac{n}{e}$ . Typically, how close are we to this number? Changing one of the  $s_i$ s can only change  $X$  by at most 1 (we can only gain or lose up to one coupon). So by the concentration inequality,

$$\Pr\left(\left|X - \frac{n}{e}\right| \geq \lambda\sqrt{n} + 1\right) \leq \Pr(|X - \mathbb{E}[X]| \geq \lambda\sqrt{n}) \leq 2e^{-\lambda^2/2},$$

where the +1 is for the approximation of  $\frac{1}{e}$  we made. So the number of coupons we miss is pretty concentrated! This would have been more difficult to solve without Azuma's inequality, because whether or not two different coupons are collected are dependent variables.

### Theorem 8.11

Let  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \in \{-1, 1\}^n$  be uniformly and independently chosen. Fix vectors  $v_1, \dots, v_n$  in some norm space (Euclidean if we'd like) such that all vectors  $|v_i| \leq 1$ . Then  $X = \|\varepsilon_1 v_1 + \dots + \varepsilon_n v_n\|$  is pretty concentrated around its mean:

$$\Pr(|X - \mathbb{E}[X]| \geq \lambda\sqrt{n}) \leq 2e^{-\lambda^2/2}.$$

Even if we can't compute the mean of this variable, we can still get concentration! Note that if the  $v_i$ s all point along the same axis, then we essentially end up with the Chernoff bound.

*Proof.* Our  $\Omega_i$ s are  $\{-1, 1\}$ , and we have a function defined as

$$f(\varepsilon_1, \dots, \varepsilon_n) = \|\varepsilon_1 v_1 + \dots + \varepsilon_n v_n\|.$$

If we change a coordinate of  $f$ , the norm can change by at most 2 by the triangle inequality, because each  $v_i$  has norm at least 1. Plugging this into Azuma, we find that

$$\Pr(|X - \mathbb{E}[X]| \geq \lambda\sqrt{n}) \leq 2e^{-\lambda^2/8}.$$

This is usually good enough (the exponent is of the right order), but with a little more care, we can change the constant in our exponent from  $\frac{1}{8}$  to  $\frac{1}{2}$ . Let's go back to the exposure martingale, and let  $Y_i$  be the expected value of our function  $f$  after having  $\varepsilon_1, \dots, \varepsilon_i$  revealed: we claim that  $|Y_i - Y_{i-1}| \leq 1$ .

Why is this the case? If we've revealed the first  $i-1$  coordinates, let  $\vec{\varepsilon}$  and  $\vec{\varepsilon}'$  be two vectors in  $\{-1, 1\}^{\varepsilon}$  differing only in the  $i$ th coordinate. Then

$$Y_{i-1}(\vec{\varepsilon}) = \frac{Y_i(\vec{\varepsilon}) + Y_i(\vec{\varepsilon}')}{2} :$$

we should average over what happens in the  $i$ th coordinate if we already know what happens in the  $i$ th coordinate. So now plugging in,

$$|Y_i(\varepsilon) - Y_{i-1}(\varepsilon)| = \frac{1}{2}|Y_i(\vec{\varepsilon}) - Y_i(\vec{\varepsilon}')| \leq \frac{1}{2} \cdot 2\|v_i\| \leq 1,$$

and now Azuma gives us the desired constant! □

## 8.4 Concentration of the chromatic number

Last time, we derived (using Janson's inequality) an estimate for the chromatic number of a random graph, and this took some work. But it turns out that we can prove concentration of the chromatic number without knowing the mean:

### Theorem 8.12

Let  $G = G(n, p)$  be a random graph. Then

$$\Pr(|\chi(G) - \mathbb{E}[\chi(G)]| \leq \lambda\sqrt{n-1}) \leq 2e^{-\lambda^2/2}.$$

To prove this, let's think about the process of finding the chromatic number as a martingale. This does not even require us knowing that  $\chi(G)$  is about  $\frac{n}{2\log_2 n}$ : proving concentration is somehow easier than finding the mean here!

*Proof.* There are many ways to expose the edges of a graph: sometimes we need to choose between edge and vertex exposure. Here, we'll do the latter.

Consider the vertex-exposure martingale. Basically, we're given the status of all edges connected to one of the first  $i$  vertices, and then we try to figure out the estimate from there. Note that  $|Z_i - Z_{i-1}| \leq 1$ : we could always just give the  $i$ th vertex a new color to preserve proper coloring, so the expected chromatic numbers can't differ by more than 1. But now we're done by applying Azuma's inequality. □

Let's try seeing what happens if we used the edge-exposure martingale instead. We have more steps: there are  $\binom{n}{2}$  edges to reveal, so we should get about  $\Theta(n)$ -size deviation. That's already not so good, since we're trying to find chromatic number (which is size  $n$ ). We can't even bound  $|Z_i - Z_{i-1}|$  any better than before!

**Remark.** *It's important to note that in general, it's not always better to use the vertex or edge exposure martingale. Instead, our method really depends on what the maximum differences are between subsequent steps.*

As a final note, can we also set up a Lipschitz function to rephrase the setting of this problem? Our random variable spaces in the vertex-exposure martingale,  $\Omega_i$ , aren't edges but batches of edges: at step  $i$ , we want  $\{0, 1\}^{i-1}$ , the edges going "left" from vertex  $i$ . So the vertex-exposure partitions our edge and exposes groups at a time: if we do the batching appropriately, we get a better gain than naively having our probability spaces just be  $\{0, 1\}$  for each edge. So the  $\Omega_i$ s are fairly general, but they must still be independent.

The idea of a tail bound is the same as a confidence interval: Azuma tells us that we can take an interval with width on the order of  $\sqrt{n}$ , and at least some constant fraction of our random graphs will give chromatic number in that interval. This might be overly generous, though: it's a major open problem to know the actual fluctuation of the chromatic number of a random graph!

## 8.5 Four-point concentration?

Interestingly, we can get even better concentration if we have sufficiently small  $p$ :

### Theorem 8.13

Let  $\alpha > \frac{5}{6}$  be a fixed constant. If  $p < n^{-\alpha}$ , then the chromatic number  $\chi(G(n, p))$  is concentrated among four values with high probability. Specifically, there exist a function  $u = u(n, p)$  such that as  $n \rightarrow \infty$ ,

$$\Pr(u \leq \chi(G(n, p)) \leq u + 3) = 1 - o(1).$$

In other words, sparser graphs are easier to estimate. Because the probability of an edge appearing here is relatively small, we can get more concentration than with our earlier calculations.

*Proof.* It suffices to show that for any fixed  $\varepsilon > 0$ , we can find a sequence  $u = u(n, p, \varepsilon)$  such that as  $n \rightarrow \infty$ ,

$$\Pr(u \leq \chi(G(n, p)) \leq u + 3) > 1 - \varepsilon - o(1).$$

Pick  $u$  to be the smallest positive integer such that  $\Pr(\chi(G(n, p)) \leq u) > \varepsilon$ . (This is deterministic, even if we may not know how to evaluate it.) Then the probability that  $\chi(G) < u$  is at most  $\varepsilon$ , and we just want to show that  $\chi(G) > u + 3$  with probability  $o(1)$ .

The next step is very clever: let  $Y = Y(G)$  be the minimum size of a subset of the vertices  $S \subset V(G)$  such that  $G - S$  may be properly colored with  $u$  colors. Basically, we color as well as we can, and  $Y$  tells us how close we are to success.

Now  $Y$  is 1-Lipschitz with respect to the vertex-exposure martingale: if we change a vertex in  $G$ , then  $Y(G)$  changes by at most 1. So by Azuma's inequality,

$$\Pr(Y \leq \mathbb{E}[Y] - \lambda\sqrt{n}) \leq e^{-\lambda^2/2},$$

$$\Pr(Y \geq \mathbb{E}[Y] + \lambda\sqrt{n}) \leq e^{-\lambda^2/2},$$

This trick will come up a lot: we'll use both the upper and lower tail separately. We don't need to know the expectation to find concentration, but we'll use the lower-tail bound to bound  $\mathbb{E}[Y]$ . With probability at least  $\varepsilon$ ,  $G$  is  $u$ -colorable. That's equivalent to saying that  $Y = 0$ , which occurs with probability

$$\varepsilon < \Pr(Y \leq \mathbb{E}[Y] - \mathbb{E}[Y]) \leq \exp\left(-\frac{\mathbb{E}[Y]^2}{2n}\right).$$

Simplifying this, this already gives us a bound (for some  $\lambda$ , which is a function of  $\epsilon$ )

$$\mathbb{E}[Y] \leq \sqrt{2 \log\left(\frac{1}{\epsilon}\right) n} = \lambda\sqrt{n},$$

which is what we should expect from a martingale of this form. Similarly, we can do an upper-tail bound to show that  $Y$  is rarely too big relative to the mean:

$$\Pr(Y \geq 2\lambda\sqrt{n}) \leq \Pr(Y \geq \mathbb{E}[Y] + \lambda\sqrt{n}) \leq e^{-\lambda^2/2} = \epsilon$$

by the definition of  $\lambda$ . So the number of uncolored vertices is not too big: by the definition of  $Y$ , we now know that with probability at least  $1 - \epsilon$ , we can color all but  $2\lambda\sqrt{n}$  vertices. Here's the key step: We'll show that with high probability, we can color the remaining vertices with just 3 colors.

**Lemma 8.14**

Fix  $\alpha > \frac{5}{6}$  as before, as well as a constant  $C$ , and let  $p < n^{-\alpha}$ . Then with high probability, every subset of size at most  $C\sqrt{n}$  vertices in  $G(n, p)$  can be properly 3-colored.

We want to union bound the bad probabilities, but we must be a bit careful here. Suppose the lemma were false for some graph  $G$  (that is, we're in one of the bad cases). Choose a minimal size  $T \subset V(G)$  that is not 3-colorable. Consider the induced subgraph  $G[T]$  (taking only the edges between the vertices in  $T$ ). This has minimum degree 3, because if there's a vertex  $x$  with  $\deg_T(x) < 3$ , then  $T - x$  is also not 3-colorable, which contradicts the minimality of  $T$ .

So  $G[T]$  has at least  $3|T|/2$  edges, and now we can just bound the probability that there exists some  $T$  (of size at least 4, since that's the only way for it to be not 3-colorable) with  $|T| \leq C\sqrt{n}$  that contains at least  $3|T|/2$  edges: union bounding, this is at most

$$\leq \sum_{t=4}^{C\sqrt{n}} \binom{n}{t} \binom{\binom{t}{2}}{3t/2} p^{3t/2}.$$

Now we just need to show that this quantity is  $o(1)$  (as  $n$  goes to  $\infty$ ): this simplifies to

$$\leq \sum_{t=4}^{C\sqrt{n}} \left(\frac{ne}{t}\right)^t \left(\frac{te}{3}\right)^{3t/2} n^{-3t\alpha/2} = \sum_{t=4}^{C\sqrt{n}} \left(O(n^{1-3\alpha/2+1/t})\right)^t = o(1)$$

if  $\alpha > \frac{5}{6}$ .

So in summary, we know that once we have all but  $C\sqrt{n}$  of the points colored with  $u$  colors with  $1 - o(1)$  probability, we have  $1 - o(1)$  probability of coloring the rest in at most 3 colors. Now just take  $\epsilon$  arbitrarily small to show the result. □

The hardest part of this proof is finding an informative random variable that is Lipschitz! It turns out that better bounds are known: we actually have two-point concentration for all  $\alpha > \frac{1}{2}$ , and the proof comes from refinements of this technique.

## 8.6 Revisiting an earlier chromatic number lemma

Remember that when we discussed Janson's inequality, we considered the following key claim from Bollobás' paper, helpful for taking out large independent sets:

### Lemma 8.15

Let  $\omega(G)$  be the number of vertices in the largest clique of  $G(n, p)$ , and let  $k_0$  be the minimum positive integer such that  $\binom{n}{k_0} 2^{-\binom{k_0}{2}} \geq 1$ . Then if  $k = k_0 - 3$  (here  $k \sim 2 \log_2 n$ ),

$$\Pr\left(\omega\left(G\left(n, \frac{1}{2}\right)\right) < k\right) = e^{-n^{2-o(1)}}.$$

(This is also Lemma 7.20.) Here's an alternative proof.

*Proof.* Let  $Y$  be the maximum number of **edge-disjoint** sets of  $k$ -cliques in  $G$ .  $Y$  is not 1-Lipschitz with respect to vertex-exposure: for example, my graph could have a bunch of cliques connected to only one point. However, it *is* 1-Lipschitz with respect to edge-exposure (since each edge can only be part of one  $k$ -clique anyway).

So now the probability that  $\omega(G) < k$  is the probability  $Y = 0$  (there are no cliques). Using the lower tail Azuma's inequality,

$$\Pr(Y = 0) = \Pr(Y \leq \mathbb{E}[Y] - \mathbb{E}[Y]) \leq \exp\left(-\frac{\mathbb{E}[Y]^2}{2\binom{n}{2}}\right).$$

It now remains to show that  $\mathbb{E}[Y]$  is large: if we can show that the expected value is  $n^{2-o(1)}$ , then lower tail estimates tell us that it is very rare for  $Y$  to be 0.

So we have this graph  $G\left(n, \frac{1}{2}\right)$ , and we're asking how many  $k$ -cliques we can pack into it. Remember the problem set: the trick is to create an auxiliary graph  $H$  whose vertices are  $k$ -cliques of  $G$ . Then two cliques  $S, T$  are adjacent in  $H$  if they overlap in at least two vertices, so they have a common edge.

Now  $Y = \alpha(H)$  is the size of the largest independent set in  $H$ : by Caro-Wei, it suffices to show that  $H$  has lots of vertices and not that many edges, since we get the convexity bound

$$\alpha(H) \geq \sum \frac{1}{1 + d(v)} \geq \frac{|V(H)|}{1 + \bar{d}}.$$

By the second moment method,  $|V(H)|$ , the number of  $k$ -cliques in  $G$ , is concentrated with high probability around its mean, which is  $\binom{n}{k} 2^{-\binom{k}{2}}$  by linearity of expectation. By definition of  $k_0$ , this is at least  $n^{3-o(1)}$ , and on the other hand, the expected number of edges in  $H$  is concentrated around  $\frac{\mathbb{E}[|V(H)|]^2 k^4}{2n^2} = n^{4+o(1)}$ . So by Caro-Wei, the expected value of  $Y$  is

$$\mathbb{E}[Y] = \mathbb{E}[\alpha(H)] \geq \mathbb{E}\left[\frac{|V(H)|^2}{|V(H)| + 2|E(H)|}\right] \geq n^{2-o(1)},$$

as desired. But here's another way to reach that conclusion, which we've seen a few times: the sampling technique!

Choose a  $q$ -random subset of  $k$ -cliques in  $G$  (each clique with probability  $q$ ). We'll just get rid of one clique from each overlapping pair to get a large  $\alpha(H)$ . We expect to get  $\mathbb{E}[q|V(H)|]$  cliques, but  $\mathbb{E}[q^2|E(H)|]$  overlapping pairs (since  $H$  is random as well). Now pick  $q$  to maximize:  $q = \frac{|V(H)|}{2|E(H)|}$ , and this means that the expected size of our independent set is at least  $\frac{\mathbb{E}[|V(H)|]^2}{2\mathbb{E}[|E(H)|]}$  and we're done.  $\square$



MIT OpenCourseWare  
<https://ocw.mit.edu>

18.218 Probabilistic Method in Combinatorics  
Spring 2019

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.